

REMARKS

The above amendment with the following remarks is submitted to be fully responsive to the Office Action of September 21, 2005. Reconsideration of this application in light of the amendment and the allowance of this application are respectfully requested.

Claims 1, 2, 4-7, 9-21 and 23-34 were pending in the present application prior to the above amendment, claims 3, 8, and 22 having been canceled. In response to the Office Action, claims 1, 7, 12, 17, 18, 21, 23, 24, 32, and 33 have been amended above. Therefore, claims 1, 2, 4-7, 9-21 and 23-34 are still pending in the present application and are believed to be in proper condition for allowance.

Referring now to the Office Action, the Examiner's withdrawal of the previous Final Office Action, and the issuance of the present non-final Office Action is noted herein. As can be appreciated by the Examiner, the present application has been extensively prosecuted since its original filing over six years ago on April 7, 1999. Whereas the Applicants acknowledge the newly assigned Examiner's desire to further prosecute the present invention, in view of the extensive prosecution history already of record, the Applicants respectfully request full and complete consideration of the present claims so that final disposition of the present case can be attained. In this regard, the independent claims of the present application have been amended above to more clearly define the present invention, and to place them in even better condition for allowance or appeal, if required.

Referring now to the substantive Office Action, the Examiner rejected all of the pending claims 1, 2, 4-7, 9-21 and 23-34 under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent No. 6,094,649 to Bowen et al. in view of U.S. Patent No. 6,144,991 to England. The Examiner asserts that with respect to independent claims 1, 18, 21, and 32, Bowen teaches most of the recited limitations in a method of extracting data of interest from a plurality of web sites, including creating a description of data of interest, an extraction pattern, associating the developed extraction pattern with the respective data of interest, receiving an extraction parameter, and obtaining of the data of interest by querying the web sites. The

Examiner explains that in his opinion, Bowen discloses search results sent to a user interface that is displayed, where HTML web pages are extracted by a web server. The Examiner further states that web indexes often points to information at numerous web sites and that the same web page may be indexed in different ways by different agents.

The Examiner acknowledges that Bowen does not teach developing an extraction pattern on output from the respective web site using a graphical user interface tool. However, the Examiner notes that Bowen does use a GUI. Moreover, the Examiner further asserts that England somehow discloses “developing an extraction pattern based on output from the respective web site using a graphical user interface tool.” The Examiner cites to Col. 2, line 65 to Col. 3, line 25 of England, and notes various operations a user can perform through the GUI of the browser. The Examiner concludes that it would be obvious to “include graphical user interface (GUI) in a browser to extract information in a pattern format as disclosed in England.” The Applicants respectfully disagree with the Examiner’s rejection and request reconsideration and withdrawal thereof.

The cited Bowen reference is directed to a method and system for supporting keyword searches of data items in a structured database, such as a relational database. In this regard, the reference discloses that selected data items are retrieved using a SQL query, and the retrieved data values are documented using a markup language such as HTML. The documents are indexed using a web crawler, the indexing agent producing an index that associates keywords with resource locators such as URLs, hot links, file paths, or distinguished names. When a user provides a keyword query to a search engine interface, the index is used to obtain a resource locator that is associated with the keyword, and used to retrieve the item’s current data from the structured database, and to generate the document containing the retrieved data. In other words, as clearly shown by the cited portions of Bowen relied upon by the Examiner, this reference merely discloses search engine technology and operation thereof which searches web page indexes for web pages that match the keyword query entered into the search engine. The cited portions of Bowen, thus disclose use of a database reader that uses SQL queries to extracts the exposed data from the database. Of course,

various search engine technologies and use of SQL in databases are well known in the art, but as explained below, this is not the subject of the recited invention.

The secondary reference to England is directed to a software system which allows a guide/specialist to interact with a client on a real-time, interactive basis over the World Wide Web. The reference discloses a guide system with a special-purpose browser displaying both locally displayable frames and remotely displayable frames, where the client system utilizes a conventional browser. The remotely displayable frames are transmitted to the client so that both the guide and the client have identical views of the remotely displayable frames. Correspondingly, the cited portions of the England reference merely disclose a customized browser which allows interaction with a client system. Of course, various browser technologies are well known in the art, but is also not the subject of the recited invention.

In the above regard, such technologies disclosed in the cited Bowen and England references do not relate to the present invention and fails to disclose, or otherwise render obvious, the present invention as claimed. As explained in the previously submitted Amendments and the Appeal Brief, the present invention discloses a method and system that facilitates extraction of data of interest from a plurality of websites. In the prior art, crawlers are created by computer programmers to retrieve information from a particular web site, for example, to extract desired information for a category of products from on-line merchants for use in an electronic catalog. However, different web sites, for example, web sites of different on-line merchants, utilize different data structures. There is no standardized structure, method or protocol, for presenting and storing information or data among different web sites that is uniformly followed by different on-line merchants. In addition, each web site generally utilizes a plurality of web pages in the web site to which a user has to navigate in order to obtain the desired data of interest regarding a product available through the web site, for example.

A crawler that is created and used to extract data from web pages of one web site generally cannot be used to extract data from other web sites due to the variations in data structure, method and/or protocol implemented by other web sites. Thus, a new crawler must be created by a computer programmer to extract data for each web

site, the creation of new crawlers being time consuming and expensive. Consequently, extracting data of interest, for example, regarding a particular product from a plurality of different web sites such as merchant web sites, can be extremely difficult, expensive, and time consuming.

The present invention provides a novel method and system for extracting data of interest from a plurality of web sites that greatly facilitates the extraction process by providing tools that can be used, even by non-programmers, to extract desired information from the plurality of web sites. (See Pg. 3, lines 9-10; Pg. 8, lines 19-23). More specifically, the present invention allows the user to generate extraction patterns directly from the output of the web site itself, such as the HTML source view of a web browser, so that other desired information can also be extracted from the web site using the generated extraction patterns. (See Pg. 3, lines 13-20; Figs. 10, 14-18 and related disclosure in Pg. 23, lines 13-19; Pg. 24, line 18-Pg. 26, line 6). As shown and described in the specification, extraction patterns of the present invention and recited in the present claims are not the same as keyword queries.

Thus, as described in the Specification of the application, the present invention can be used by an individual such as a programmer, or even a non-programmer, to generate extraction patterns easily based on the output of the web site, such as the HTML source code, of the web site itself. A value can then be used in conjunction with the developed extraction pattern to extract different data of interest from the particular web site. Correspondingly, the present invention allows facilitated extraction of desired data of interest from a plurality of web sites in a rapid, cost effective manner, without requiring a programmer to create a new and different crawler for each web site from which data of interest is desired. Example implementation of the present invention and development of extraction patterns are most clearly shown in Figures 14 to 17, and the corresponding portions of the specification describing these figures in Page 24, line 18 to Page 25, line 24. The Examiner is respectfully referred to the cited portions of the specification of the present application to view example implementations of the present invention as well as examples of the recited extraction patterns. These examples were also clearly explained in the Amendment filed September 3, 2003.

Clearly, the Examiner's rejection of the pending claims is improper. First, there is no motivation established in either of these references to combine them, and to modify the system of Bowen using the teachings of the cited England reference, as suggested by the Examiner. This is expected since these cited references are directed to two different technological fields that are not related, Bowen being directed to system for supporting keyword searches of data items in a structured database (such as a relational database), while England is directed to a guide system with a special-purpose browser displaying both locally displayable frames and remotely displayable frames. The Examiner fails to identify any teachings in either references to combine them in any manner. Because there is no teachings or suggestions in either of the cited references to combine the references in the manner suggested by the Examiner, the Applicants respectfully contend that the Examiner has failed to properly establish a *prima facie* case of obviousness. Correspondingly, the withdrawal of the present rejection is respectfully requested.

Secondly, even if there was motivation to combine these references in the manner suggested by the Examiner, they fail to result in the present invention as claimed. In particular, whereas search engines such as that described in Bowen are adapted to search for, and retrieve, web pages that match a query that is specifically entered into the search engine, they cannot be used to extract data for which specific parameters are not entered as a query since there will be no matching of data for retrieval by the search engine. Thus, for example, whereas the extraction patterns developed in accordance with the present invention can be utilized to extract data associated with all books from a web site since such data will likely be structured in a similar manner in a particular web site, mere search engines such as that described in Bowen cannot perform such a function. Instead, the particulars for each book, such as a title, author, or other parameter would be required in order to retrieve information about books from a web site. As previously noted, extraction patterns of the present invention and recited in the present claims are not the same, or even equivalent to, keyword queries. Thus, even if the system disclosed in Bowen was modified to have a customized graphical user interface as disclosed in the cited England reference, such combination would still fail to result in a system as claimed that allows developing of

an extraction pattern based on output from the respective web pages. Correspondingly, the withdrawal of this rejection based on Bowen and England is also requested for the above reason as well.

However, to expedite prosecution in the present case in view of the long prosecution history, and reassignment of this case to a new Examiner by the USPTO, the independent claims 1, 18, 21, and 32 have been amended to more clearly define the present invention. In particular, these claims have been amended to specifically recite developing an extraction pattern from a web page output from the respective website, and that the extraction pattern is adapted to extract information from a plurality of web pages of the respective web site. Clearly, the cited reference, either alone or in combination, fails to disclose, teach, or otherwise suggest development of an extraction pattern from a web page output of a website. In addition, the cited references or combination thereof fails to disclose, or otherwise suggest that the extraction pattern is adapted to extract information from a plurality of web pages of the respective web site. Therefore, the withdrawal of this obviousness rejection, and the allowance of independent claims 1, 18, 21, and 32 are respectfully requested.

Referring again to the Office Action, dependent claims 2, 5, 7, and 20 were rejected based on England. However, this rejection is believed to be improper in view of the discussions above, and also rendered moot in view of their dependency on independent claims 1 or 18. In addition, England clearly does not disclose or suggest a plurality of predefined extraction patterns as recited in claim 5, or the functions of the graphical user interface tool as recited in claim 7. Correspondingly, the withdrawal of this rejection and allowance of these claims are respectfully requested.

Referring to the Office Action, dependent claims 4 and 9-17 were rejected based on Bowen. However, this rejection is believed to be improper in view of the discussions above, and also rendered moot in view of their dependency on independent claim 1. Moreover, in contrast to the Examiner's assertion, Bowen does not disclose or suggest application of an extraction pattern to the output of the website displayed in a source view as specifically recited in claim 4. The cited portion of the Bowen reference relied upon by the Examiner merely describes operation of indexing technology that utilizes SQL to search for, and retrieve, documents that match the

query. The reference fails to disclose application of an extraction pattern to the output of a web site, much less an output displayed in a source view. Furthermore, just because different views may be possible in prior art systems, this does not mean that an extraction pattern is applied to a particular view as recited. Clearly, this summary rejection of claim 4 is improper.

In addition, the Examiner's basis for rejecting the dependent claims 9-17 is unclear, these claims reciting features that are not disclosed or suggested by the Bowen reference. In particular, the cited Bowen reference fails to disclose or suggest selection of an extraction command from a list of extraction commands as recited in claim 9, the particular extraction commands recited in claims 10, 11, 16 and 17, or the application of a test condition having a logical test as recited in claims 12-15. Therefore, the withdrawal of this rejection and the allowance of these claims are also respectfully requested.

Referring again to the Office Action, claims 6, 25, and 26 were also rejected based on Bowen. However, this rejection is believed to be rendered moot in view of the above, these claims being ultimately dependent on independent claim 1 which is believed to be in proper condition for allowance. Moreover, the basis for this rejection is wholly unclear in that claim 6 requires a GUI interface tool with a plurality of predefined extraction patterns, at least one of which matches a particular type of information. The cited portion of the Bowen reference merely discloses various types of documents that may include links, for example. Of course, such documents are known, but the Bowen reference does not disclose predefined extraction patterns of the present invention that perform particular function as recited in claim 6. Thus, the withdrawal of this rejection and the allowance of these claims are also respectfully requested.

Claim 9 has been rejected based on Bowen, claim 9 reciting developing of the extraction pattern including receiving a selection of an extraction command from a predetermined list. This rejection is believed to be moot in view of the fact that it is dependent on independent claim 1 which is believed to be in proper condition for allowance. In addition, the basis for this rejection is not understood since the cited portions of the Bowen reference does not disclose an extraction pattern or a

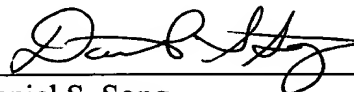
predetermined list of commands at all, but instead discusses hardware of a computer and various programmable languages. Thus, the Applicants contend that this rejection is improper and should be withdrawn.

Referring again to the Office Action, claims 23, 24, 27, 28 and 33 were also rejected under Bowen. The portions of Bowen cited by the Examiner only relates to keyword searching and use of wildcards in such keyword searches. The cited portions do not relate at all to the recited extraction patterns, or such patterns with various expressions that perform the functions of extraction. In this regard, as previously noted, extraction patterns are not the same as, or equivalent to, keyword queries. Thus, the withdrawal of this rejection and the allowance of claims 23, 24, 27, 28 and 33 are respectfully requested.

Finally, referring again to the Office Action, claims 29-31 and 34 were also rejected based on Bowen. However, this rejection is believed to be rendered moot in view of the above comments, and the fact that these claims are ultimately dependent on an allowable independent claim. Therefore, the withdrawal of this rejection and the allowance of these claims are further requested.

In view of the foregoing, it is submitted that the present application is in condition for allowance and a notice to that effect is respectfully requested. However, if the Examiner deems that any issue remains after considering this response, he is invited to call the undersigned to expedite the prosecution and work out any such issue by telephone.

Respectfully submitted,



Daniel S. Song

Registration No. 43,143

NIXON PEABODY LLP
401 9th Street, N.W., Suite 900
Washington, D.C. 20004-2128
(202) 585-8000
(202) 585-8080 (Fax)
Customer No. 22204